

Roll No

CS-503 (A) (GS)**B.Tech. V Semester**

Examination, November 2023

Grading System (GS)**Data Analytics****Time : Three Hours****Maximum Marks : 70****Note:** i) Attempt any five questions.

किन्हीं पाँच प्रश्नों को हल कीजिए।

ii) All questions carries equal marks.

सभी प्रश्नों के समान अंक हैं।

iii) In case of any doubt of dispute the english version question should be treated as final.

किसी भी प्रकार के संदेह अथवा विवाद की स्थिति में अंग्रेजी भाषा के प्रश्न को अंतिम माना जायेगा।

1. a) Suppose the weights of 800 male students are normally distributed with 28.8 kg and SD of 2.06 kg. Find the number of students whose weights are

7

i) Between 28.4 kg and 30.4 kg

ii) More than 31.3 kg

मान लीजिए कि 800 छात्रों का वजन सामान्य रूप से 28.8 किग्रा और SD 2.06 किग्रा के साथ वितरित किया जाता है। उन छात्रों की संख्या ज्ञात कीजिए जिनका भार है

i) 28.4 किग्रा और 30.4 किग्रा के बीच

ii) 31.3 किग्रा से अधिक

b) A random variable has the following probability function:

7

x	0	1	2	3	4	5	6	7
$P(x)$	0	K	2K	2K	3K	K^2	$2K^2$	$7K^2 + K$

Determine:

i) k

ii) mean

iii) variance

एक यादृच्छिक चर में निम्नलिखित प्रायिकता फलन होता है:

x	0	1	2	3	4	5	6	7
$P(x)$	0	K	2K	2K	3K	K^2	$2K^2$	$7K^2 + K$

निर्धारित करें:

i) k

ii) mean

iii) विचरण

2. a) A sales tax officer has reported that the average sales of the 500 businesses that he has to deal with during a year is Rs. 36,000 with a standard deviation of Rs. 10,000. Assuming that the sales in these businesses are normally distributed, find:

7

i) The number of business as the sales of which are greater than Rs. 40,000.

ii) The percentage of business the sales of which are likely to range between Rs. 30,000 and Rs. 40,000.

एक बिक्री कर अधिकारी ने बताया है कि एक वर्ष के दौरान उसे जिन 500 व्यवसायों से निपटना है, उनकी औसत बिक्री रु. 36,000, रुपये 10,000 के मानक विचलन के साथ है। यह मानते हुए कि इन व्यवसायों में बिक्री सामान्य रूप से वितरित की जाती है, खोजें:

i) उन व्यवसायों की संख्या जिनकी बिक्री 40,000 रुपये से अधिक है।

ii) व्यवसाय का प्रतिशत जिसकी बिक्री 30,000 रुपये और 40,000 रुपये के बीच होने की संभावना है।

- b) Discuss the trends in big data generation and acquisition. 7

बड़े डाटा के निर्माण और अधिग्रहण के रुझानों पर चर्चा करें।

3. Explain the following: 14

- Predictive analytics
- Inter-and Trans-firewall analytics
- Information management
- Crowd sourcing analytics

निम्नलिखित की व्याख्या करें:

- भविष्य बतानेवाला एनालिटिक्स
- इंटर-और ट्रांस-फ़ायरवॉल एनालिटिक्स
- सूचना प्रबंधन
- क्राउड सोर्सिंग एनालिटिक्स

4. a) With an example, explain the term social media analytics. 7

उदाहरण के साथ, सोशल मीडिया एनालिटिक्स शब्द की व्याख्या करें।

- b) What are the various stages in big data analytics life cycle? Illustrate with a figure, explaining each of them. 7

बिग डाटा एनालिटिक्स जीवन चक्र में विभिन्न चरण क्या हैं? उनमें से प्रत्येक को समझाते हुए एक आकृति के साथ चित्रित करें।

5. a) Brief about the main component of Map Reduce. 7

मैप रिड्यूस के मुख्य घटक के बारे में संक्षेप में बताइए।

- b) What is Hadoop? Describe the role of Hadoop in big data analysis. Also explain core components of Hadoop. 7

हडूप क्या है? बड़े डाटा विश्लेषण में हडूप की भूमिका का वर्णन करें। Hadoop के मुख्य घटकों के बारे में भी बताइए।

6. a) Describe the structure of HDFS in a Hadoop ecosystem using a diagram. 7

हडूप पारिस्थितिकी तंत्र में HDFS की संरचना का वर्णन एक आरेख का उपयोग करके करें।

- b) Why is finding similar items important in Big Data? Illustrate using two example applications. 7

बिग डाटा में समान आइटम ढूँढना क्यों महत्वपूर्ण है? दो उदाहरण अनुप्रयोगों का उपयोग करके समझाइए।

7. a) Why to choose Hadoop for processing Big Data in detail and explain the concept of distributed and parallel computing challenges? 7

बड़े डाटा को विस्तार से संसाधित करने के लिए हडूप का चयन क्यों करें और वितरित और समांतर कंप्यूटिंग चुनौतियों की अवधारणा की व्याख्या करें?

- b) Explain in detail the interacting process with Hadoop Ecosystem. List out various big data processing technologies. 7

हडूप पारिस्थितिकी तंत्र के साथ परस्पर क्रिया प्रक्रिया को विस्तार से समझाइए। विभिन्न बड़ी डाटा प्रोसेसिंग तकनीकों की सूची बनाइए।

8. a) Explain Pig Data Model in detail and discuss how it will help for effective data flow? 7

पिग डाटा मॉडल की विस्तार से व्याख्या करें और चर्चा करें कि प्रभावी डाटा प्रवाह के लिए यह कैसे मदद करेगा?

- b) Draw and explain architecture of APACHE HIVE. Explain various data insertion techniques in HIVE with example. 7

APACHE HIVE की वास्तुकला को चित्रित करें और समझाइए। HIVE में विभिन्न डाटा सम्मिलन तकनीकों को उदाहरण सहित समझाइए।
